The Relationship of Word Use and the COVID-19 pandemic in Bloggers with Mental Disorders

Linde, Carolin

The Relationship of Word Use and the COVID-19 pandemic in Bloggers with Mental Disorders

Name author: Carolin S. Linde

ID number: S2899337

Name supervisor: Marc Molendijk

Master's degree program: Clinical Psychology

Institute of Psychology

Universiteit Leiden

Date: 30.04.2021

Abstract

Word use is proven to be an indicator for psychological health and to be related to stress, especially the word categories of *I-use* and *verb tense* are related to psychiatric symptoms. This study examines the relationship of I-use and verb tense in bloggers with mental disorders and healthy controls prior and during COVID-19 (N=25). 1850 blogposts were extracted from online platforms and analyzed by text analysis software. Using a multiple-case report with quantitative analyses over cases, the changes in the use of *I*, *Future Orientation* and *Past Orientation* before and during COVID-19 were assessed. Results indicated a significant main effect of patient status on word use. The clinical group used more first-person singular pronouns and less future tense, relative to the control group ($p<.01$, $d=-.53$; $p<.01$, $d=.39$). Random effects meta-analysis showed no change in any of the word categories from pre COVID to COVID. A difference in the change of the use of *Focus Past* was found between the clinical group and the control group ($Z=2.23$, $p<.01$), with a larger reduction in the clinical group. There was no significant difference in word use between blogs dealing about COVID (n=87) and blogs not dealing about COVID (n=826). This study shows that word use is related to mental health and that COVID-19 does not seem to influence word use. It was one of the first to examine the effect of COVID-19 on word use reflecting mental health and underlies the importance of conducting further research regarding this relationship.

Keywords: COVID-19, word use, mental health, LIWC

What makes us humans unique from other species is our communication with a very sophisticated and detailed use of language. People speak an average of at least 16.000 words per day and thus, the use of language is a very dominant process in our everyday life, creating a very important field of research (Mehl et al., 2007). In our daily life, it is easily recognizable that individuals differ in how they talk and write. A study by Pennebaker and King (1999) supports the hypothesis that word use is related to personality style by revealing that individuals differ in a stable manner in their use of certain word categories. In their first analysis they included daily diaries from 15 substance abuse inpatients, daily writing assignments of 35 students and the abstracts of different journal articles written by 40 social psychologists. In a follow-up analysis including the most common word categories, essays from 838 students were analyzed. The results revealed that especially the use of emotional

language, including positive and negative emotion words, the use of long words and the use of verb tenses differed within the investigated samples and that this difference was stable within individuals also when writing about different topics. This study created a building block for a new, more objective approach to study personality, namely by analyzing written materials with regard to its linguistic properties.

Using this theoretical foundation that word use differs between individuals, more specific populations of interest, individuals diagnosed with mental disorders, were studied. With a sample of 27 psychiatric outpatients and 17 nonclinical controls, Jungheanel et al. (2008) examined linguistic differences between an outpatient group and a nonclinical control group. In standardized writing assignments, the participants were instructed to write in a detailed manner about a personal topic, for example about their family and their thoughts and emotions regarding this topic. Those assignments were analyzed, and the results showed amongst others that the psychiatric patients used fewer words reflecting optimism, cognitive mechanisms, future orientation and communication. In line with these results, Brockmeyer et al. (2015) found that during negative memory recall, in patients with anorexia nervosa being comorbid depressed and anxious, the use of first-person singular pronouns was positively related to symptoms of depression and anxiety, which was also found during a follow up study examining non-depressed individuals. Adding to these outcomes, a study conducted by Arntz et al. (2012) revealed first, that essays written by patients diagnosed with a personality disorder differed amongst others in the proportional word use of emotion words, verb tense and first-person singular pronouns compared to essay written by healthy controls. Secondly, they showed that this difference became smaller during the course of therapy and additionally, that the reduction of negative emotion words was associated with a better treatment outcome.

To conclude, word use differs between individuals and there are significant differences in the use of different word categories when comparing individuals with mental disorders and healthy control samples. The above presented studies amongst others focused on the differences in the use of first-person singular pronouns and verb tenses. In the following, these word categories will be classified into broader psychological concepts. The significant difference in the use of first-person singular pronouns as observed by Brockmeyer et al. (2015) and Arntz et al. (2012) is in line with the *self-awareness theory of reactive depression* as proposed by Pyszczynski and Greenberg (1987). This theory suggests that the onset of depression is precipitated by any kind of losses triggering the experience of negative

affect. A self-regulatory cycle is activated in which the individual focuses on regaining the loss and is not able to disengage from this cycle. This cycle leads to an excessive self-focus especially regarding negative events while avoiding the self-focus on positive events. These results are supported by those obtained by Harvey et al. (2004) proposing a transdiagnostic approach to psychopathology, meaning that there are several common mechanisms underlying many psychological disorders. Amongst others they suggest that several attentional processes, including self-focused attention underlies different forms of psychopathology. This includes for example social anxiety in which individuals become self-focused on thoughts, feelings, bodily sensations, first before the feared situation and second after leaving the situation (Clark, 2012). Thus, different mental disorders seem to be characterized by an excessive self-focus which is reflected in the use of first-person singular pronouns.

It has also been shown that verb tense is differently used in individuals with mental disorders compared to healthy controls (Pennebaker et al., 1999; Junheanel et al., 2008; Arntz et al., 2012). This finding could be further explained and classified with the concepts of rumination and worrying. Rumination is a depressive symptom concerning repetitive, self-focused thinking patterns regarding *past* events, most often about losses or past failure. Whereas worrying is a symptom of anxiety disorders in which individuals experience streams of thoughts concerning negative events in the *future* (Hur et al., 2017). Hence, the differently used verb tenses in psychiatric patients compared to healthy controls might be explained by the cognitive mechanisms of rumination (focusing on the past) and worrying (focusing on the future). To sum up, the presented studies show that word use is associated with psychological health which can be especially seen in the use of verb tense and first-person singular pronouns. This difference in word use between clinical samples and healthy control samples can be explained by broader concepts including self-focus, worrying and rumination.

The Role of Stress

Somewhat contradictory to the claim that people's individual linguistic style is reliable across time and situations, is the finding of Cohn et al. (2004) showing that after the occurrence of a major stressor, the use of several word categories changed in the diary entries of 1084 US citizens. In their study, diary entries of an online platform were downloaded and analyzed two months prior and two months after the September-11 terror attacks on the twin-towers of the

Word Trade Center in the United States. Shortly after the attacks, the use of words reflecting negative emotions, psychological distance, social- and cognitive engagement words increased, whereas this returned to baseline during the following six weeks. Further, the preoccupation with September-11 and its relationship with word use was measured with 27 target words rated by independent judges as specifically related to the September-11 attacks and its aftermaths. It was found that the relationship of the September-11 attacks as stressor and the use of different word categories was mediated by the preoccupation with these attacks, meaning that individuals who were highly preoccupied showed stronger effects regarding the change in word use. Related to these results are the findings of Pennebaker and Lay (2002) who analyzed 35 press conferences of Rudolph Giuliani, the mayor of New York City from 1994 to 2001. The researchers investigated the relationship of language use during crisis, focusing on a personal crisis of the politician, in which he was diagnosed with prostate cancer, withdrew from the senate race and separated from his wife, and on a more general crisis after the September-11 attacks. Results indicated that during these crises, the politician used amongst others significantly more first-person singular pronouns, became more future oriented, expressed more emotions and was cognitively more engaged. Concluding, the experience of crises and stressors changes the use of certain word categories on short-term and this is mediated by the preoccupation with the stressor.

To define the concepts of *stressors* more concretely, it is first important to distinguish between physical and psychological stressors. The present study focusses on *psychological stressors* which can be defined as "social and physical environmental circumstances that challenge the adaptive capabilities and resources of an organism" (Monroe & Slavich, 2016, p.109). An example of such a stressor is the SARS-CoV-2 causing the COVID-19 disease which was declared as a worldwide pandemic in March 2020 by the World Health Organization (WHO, 2020). According the WHO (2020), this disease causes mild to moderate respiratory illness symptoms in people having no pre-existing diseases. For older people or people having chronic respiratory diseases, the coronavirus can develop more seriously. In many countries this pandemic led to restrictions like school and kinder garden closings, a prohibition for larger celebrations, restrictions in the number people to meet with or whole lockdowns in which everything had to close except system relevant institutions like supermarkets. There are already a few studies showing the impact this pandemic has on people's mental health. The results of a review on 19 articles about the effect of COVID-19

on mental health in the general population by Xiong et al. (2020) revealed that during the COVID-19 pandemic higher prevalence rates of anxiety, depression, post-traumatic stress disorder, psychological distress, and stress were reported. Further, the results suggest several risk factors associated with the experience of these symptoms including being female, being younger than 40 years, having a physical or mental illness, being unemployed or a student and being exposed to media or news reporting about the pandemic. A study conducted by Alonzi et al. (2020) investigated a sample of 616 young adults during the COVID-19 pandemic who completed the short form of the Patient-Reported Outcomes Information System screening for depression and anxiety symptoms. The results showed that individuals with pre-existing mental or physical disorders self-report higher depression and anxiety symptoms during the COVID-19 pandemic than individuals without pre-existing illnesses. Next to that, the results indicated that females revealed higher levels of those symptoms compared to males, which is in accordance with the results obtained by Xiong et al. (2020). Investigating a more specific sample, namely 402 COVID-19 survivors, Mazza et al. (2020) screened for different psychiatric symptoms making use of clinical interviews and self-report questionnaires. The results revealed that 56% of their sample scored in at least one clinical dimension in the pathological range, including post-traumatic stress disorder, depression, anxiety, insomnia and obsessive-compulsive symptoms. Additionally, it was shown that compared to males, females suffer more from both depression and anxiety and that individuals who already had a mental disorder before becoming infected with COVID-19, increased on most of the psychiatric measures. Concluding, the current COVID-19 pandemic affects people's mental health and especially anxiety, depression and post-traumatic stress disorder symptoms are increasing. Several risk factors have already been identified including female gender, young age (<40) and pre-existing physical or mental disorders.

There is very little research about how COVID-19 affects word use but for example, a study by Su et al. (2020) extracted posts from online platforms in Wuhan and Italy two weeks before and two weeks after the local lockdowns and analyzed those with regard to the change in different word categories using text analysis software. The results of the changes in word use of the platform in Wuhan (*Weibo*) showed significant changes in 16 out of 39 word categories, including amongst others first-person singular and plural pronouns, emotion words, time orientation and religion words. The posts from the Italian platform (*Twitter*) showed significant changes in ten word categories, including personal concerns words and

cognitive mechanism words. A related study by Abdo et al. (2020) extracted over 900.000 English tweets from January to April 2020 to measure change in word use over 14 weeks. Their results indicated amongst others an increase in the use of the word categories *negative emotion, fear, anxiety* and *sadness.* Thus, although there is not much research on how COVID-19 is related to word use, there are already some studies showing that the pandemic changes the use of certain word categories being similar to those changing in situations of stress or crises as observed by Cohn et al. (2004) and Pennebaker and Lay (2002).

To conclude, the presented data supports the hypothesis that psychological health is reflected in people's word use and that the experience of a stressor and the preoccupation with that, influences this relationship. However, there is a deficiency of studies analyzing broader language products over a longer period of time and specifically the exploration of the relationship of stressors and word use is not investigated sufficiently. The current COVID-19 pandemic is an acute stressor whose influence on word use and psychological health still remains under-researched.

Research Aim and Implications

The purpose of this study is to investigate the relationship of the COVID-19 pandemic as psychological stressor and the word use of bloggers, specifically the word categories of verb tense and use of first-person singular pronouns. Another objective is to examine whether these variables are related differently between bloggers who suffer from a mental disorder and healthy control bloggers. This will have several practical implementations. First, the insight regarding the effects the COVID-19 pandemic has on mental health will help to specify the psychological needs of people with- and without mental disorders during this pandemic and can help to develop interventions satisfying these needs. Another practical implication of the present study is that it uses a different method to measure mental health, namely an implicit and more objective one. This measure expands the field of research regarding the measurement of people's mental health moving from more explicit questionnaires and interviews to a more implicit measurement which might have a positive influence on the objectivity of the analysis of mental health.

Hypotheses

Based on the above elaborated results of Cohn et al. (2004), Pennebaker et al. (2002), Su et al. (2020) and Abdo et al. (2020) the main hypothesis of the present study is that blogposts published after COVID-19 emerged (January to June 2020) will include significantly more first-person singular pronouns, more past-tense verbs and more future-tense verbs compared to the time period before COVID-19 emerged (January to June 2019). Further, it is hypothesized that compared to healthy control bloggers, blogposts written by individuals with mental disorders will show a significantly higher change concerning these word categories in accordance with the results of Jungheanel et al. (2008), Brockmeyer et al. (2015) and Arntz et al. (2012). Based on the study of Cohn et al. (2004), the third hypothesis is that the relation of word use and COVID-19 is mediated by the preoccupation with COVID-19. Hence, it is hypothesized that blogposts dealing about COVID-19 will show a significantly stronger effect regarding the change in the use of first-person singular words and verb tense compared to blogposts about different topics.

Methods

Research Design

This study presents a multiple-case report with quantitative analyses over cases. It has an observational retrospective design with *Time* (qualitative: Pre COVID/January to June 2019, COVID/January to June 2020) as non-experimental within-subject factor. *Patient Status* (qualitative: clinical, control) and *Topic Blogpost* (qualitative: COVID, not COVID) were included as non-experimental between-subject factors. The dependent variables in this design were proportional change in use of future tense (quantitative: -1-1), proportional change in use of past tense (quantitative: -1-1), proportional change in use of first-person singular pronouns (quantitative: -1-1).

Participants

The target group of this study were bloggers publishing written online blogs on the internet. Inclusion criteria were to write either in English, German, Dutch or Spanish. Further, a sufficient number of blogposts (>5) had to be available per timeframe of interest (January to June 2019 and January to June 2020). A specific inclusion criterium for the clinical group was to be diagnosed with or give a self-statement of having a mental disorder, whereas this was an

exclusion criterium for the control group. The participants were recruited by searching for blogs fulfilling the in- and exclusion criteria on the internet. The participants of the control group were matched to participants in the clinical group on the variables *age*, *language*, *gender* and *region of residence*. Blogs included were nine for the clinical group and nine for the healthy control group. Two blogs have more than one author creating a total sample of *N*=25. The age of participants ranged from 20 to 44 years, with a mean age of 27.95 (*SD*=6.40). The age of 13 participants was unknown. All included participants were female.

Measures

The dependent variables were generated by using the Linguistic Inquiry and Word Count (LIWC), a text analysis program developed by Pennebaker, Francis, Boyd & Booth (2015a) to create an objective software analysing written text materials with regard to their linguistic components. Psychometric properties regarding the reliability and validity of LIWC were established by Pennebaker, Boyd, Jordan and Blackburn (2015b). The program is available in many different languages and examines each word of a text document by assigning them to over 80 different linguistic word categories and subcategories. After that, a report is constructed showing the percentage of the use of each category as a function of the total number of words in the text. The categories are diverse, including linguistic categories like *prepositions, verb tenses* and psychological categories like *positive and negative emotion words* or *words referring to insight*. The categories are sorted hierarchically, and words can also overlap in different categories, for example the word *thought* will be assigned to the word category of *past tense words* and to the category of *cognitive insight.* The present study will focus on the specific word categories of *verb tense* (future and past orientation) and *first-person singular pronouns.*

First-person singular pronouns

The word category includes all first-person singular pronouns like “I, me, mine” and has a corrected internal consistency of α= .81. A higher score in this word category reflects a higher self-focus (Pennebaker et al. 2015b).

Verb tense

This linguistic word category is divided in three subcategories being future tense (corrected α= .64), present tense (corrected α= .66) and past tense (corrected α= .68). Examples of these words are "will be, go, said". Higher scores in the future tense reflect a higher future orientation, higher scores in the present tense reflect a higher orientation in the present and higher scores in the past tense display a higher orientation in the past (Pennebaker et al. 2015b).

Procedure

Blogs were searched online in different search engines like *google*, were checked for in- and exclusion criteria and healthy control blogs were matched to clinical blogs. Afterwards the single blogposts were converted into word files whereafter those were entered in the LIWC program. Lastly, the reports of the LIWC program giving proportions for the use of each word category were transformed to an excel format to be capable for the statistical analysis in SPSS and Comprehensive Meta-Analysis.

Ethics

An ethical approval was not necessary as the data was extracted from existing online blogposts available to public. Data is be reported anonymously.

Statistical Analysis

Data was prepared and analyzed with the IBM SPSS 26 package and Comprehensive Meta-Analysis (Borenstein, Hedges, Higgins, & Rothstein, 2011). Descriptive statistics were calculated overall time frames and for change in the word categories *I*, *Focus Past* and *Focus Future* within blogger. Spearman's rho coefficients analysis was conducted with a significance level of α= 0.05 to test associations between the outcome variables. Main effects of patient status were assessed by means of an independent *t*-test with patients and controls paired for language. The descriptive statistics were fed into CMA where overall within-subjects effects were calculated by means of random effects meta-analysis with a confidence level of 95%. Effect moderation by patient status and language was tested for in moderation analysis. Cohen's *d* was used to describe the standardized differences between means. For interpretation purposes, the benchmarks suggested by Cohen (2013) were used, being: small

(*d*=0.2), medium (*d*=0.5) and large (*d*=0.8). Main effects of patient status and language were assessed though moderator analyses within CMA. An independent *t*-test on the data of 2020 was conducted to analyze whether *Topic of Blog* (qualitative: COVID, not COVID) was associated with mean differences in *I*, *Focus Past* and *Focus Future* use. The significance level was α= 0.05.

Results

Table 1 displays a summary of the analyzed sample characteristics, including an anonymous ID number, the language of the blog, the disorder or psychological symptoms the clinical group experienced and the age of the blogger including means and standard deviations.

Table 1

Sample Characteristics

Clinical Group				**Control Group**		
Blogger	**Language**	**Disorder/ Psychological Symptoms**	**Age**	**Blogger**	**Language**	**Age**
ID 3	Dutch	Depression, PTSD, Autism	21	ID 1	Dutch	26
ID 4	Dutch	Anxiety, high sensitive person	20	ID 2	Dutch	25
ID 5	English	Social Anxiety	44	ID 6	English	/
ID 9	English	Anxiety	26	ID 7	English	/
ID 12	English	Depression, Anxiety	/	ID 8	English	32
ID 13	English	Depression	28	ID 10	English	24
ID 14	English	Social Anxiety	/	ID 11	English	26
ID 15	German	Panic Disorder, Agoraphobia	33	ID 17	German	/
ID 16	German	Borderline Personality Disorder	30	ID 18	German	/
			M=28,86, *SD*=8.13			*M*=26.6, *SD*=3.13

The descriptive statistics for word categories *I*, *Focus Past* and *Focus Future* within blogger giving means, standard deviations and number of before (January to June 2019) and during COVID (January to June 2020) are displayed in the Appendix. In total, 1850 blogposts

were analyzed, 937 of those were published before COVID emerged (January to June 2019) and 913 blogposts were published during the COVID pandemic (January to June 2020).

Spearman's rho Coefficients Analysis

Correlations between all outcome variables are presented in Table 2. There was a positive correlation of *Focus Future* and *I-Use* ($p<.01$). Further, the variables *Focus Past* and *I-use* were positive correlated ($p<.01$) and results indicated a positive correlation of *Focus Past* and *Focus Future* ($p<.01$).

Table 2

Spearman's rho correlation coefficients

	I-Use	Focus Future	Focus Past
I-Use	1.00		
Focus Future	.16**	1.00	
Focus Past	.53**	.36**	1.00

Note. The significance level was α= 0.05.

** Significant correlation at the 0.01 level

Between-Subjects Effects

The results of the between-subjects effects are summarized in Table 3. There was a main effect of *Patient Status* (clinical group, control group) regarding *I-use*, controlling for language. Further, a main effect of *Patient Status* (clinical group, control group) regarding the *Future Orientation*, controlling for language was found. Results indicated no main effect of *Patient Status* (clinical group, control group) on *Past Orientation*, controlling for language.

Table 3

Between-Subjects Effects Patient Status

Patient Status	Mean	SD	n	Cohen's d	95% CI	p-value
			I-Use			
Clinical Group	5.54	2.70	707	-0.53	-0.67, -0.40	<.01
Control Group	4.01	3.06	1143			
			Focus Future			
Clinical Group	1.51	1.11	707	0.39	0.32, 0.46	<.01
Control Group	1.98	1.28	1143			
			Focus Past			
Clinical Group	3.78	2.06	707	0.01	-0.09, 0.10	*p*=.83
Control Group	3.82	2.79	1143			

Note. CI=Confidence Interval. A positive effect size indicates that the clinical group used fewer words of the word category relative to the control group.

Within-Subject Effects

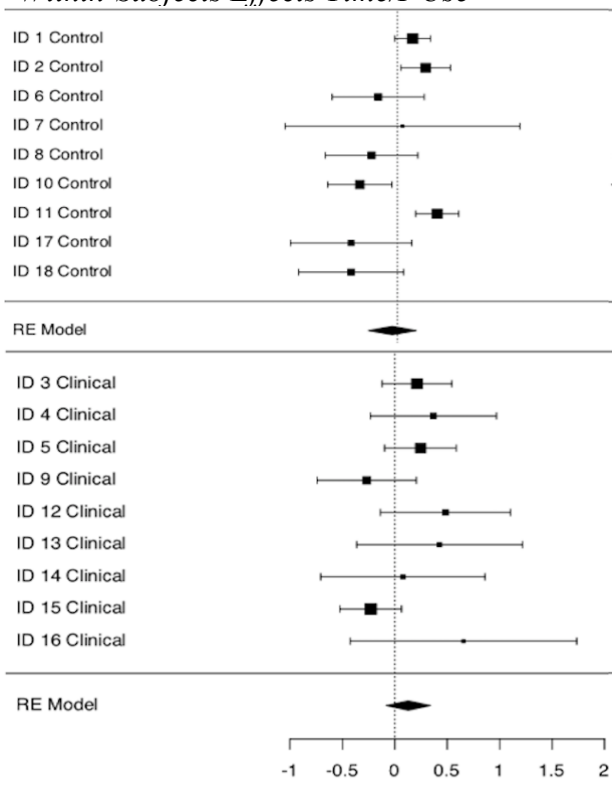
In the following the results of the random effects meta-analysis are presented which analyzed the change in word use from January until June 2019 (Pre COVID) to January until June 2020 (COVID) per subject. The forests plots display the effect of *Time* (Jan-Jun 2019/pre COVID, Jan-Jun 2020/COVID) on the use of the analyzed word categories. On the vertical axis the bloggers are represented, separated by control- and clinical group. The vertical line is the "line of null effect" being set at null, indicating no association between the variables. On the horizontal axis, point estimates are represented by solid squares and horizontal lines indicate the 95% confidence interval of the individual estimate. The size of the squares corresponds to the weight this blogger had in the meta-analysis, determined by the number of blogposts included. The diamond represents the summary statistics with its lengths indicating the confidence interval.

I-use

Figure 1 presents the within-subjects effect of *Time* (Jan-Jun 2019/pre COVID, Jan-Jun 2020/COVID) on *I-use* for each subject and a summarized effect for the control- and clinical group. The results indicated no effect of *Time* on *I-use*. Further, there was no significant difference in the change of word use between the clinical group and the control group (Z=1.11, p=0.14). The effects did not differ within language.

Figure 1

Within-Subjects Effects Time/I-Use

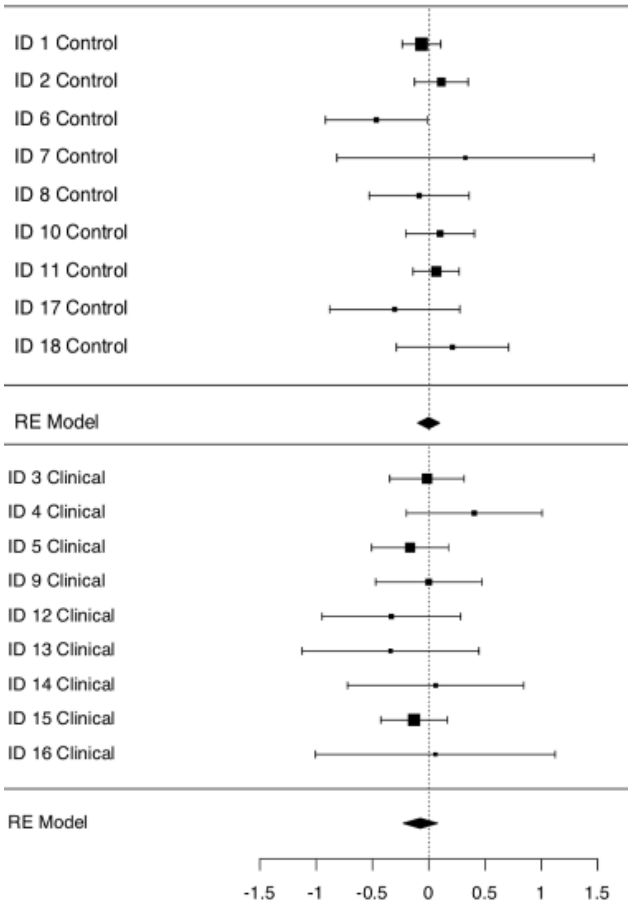


Focus Future

Figure 2 displays the within-subjects effect of *Time* (Jan-Jun 2019/ pre COVID, Jan-Jun 2020/COVID) on the use of the word category *Focus Future* and a summarized effect for the control- and clinical group. There was no effect of *Time* on the use of the word category *Focus Future*. Further, there was no significant difference in the change of word use between the clinical group and the control group (Z=0.66, p=0.89). The effects did not differ within language.

Figure 2

Within-Subjects Effects Time/Focus Future

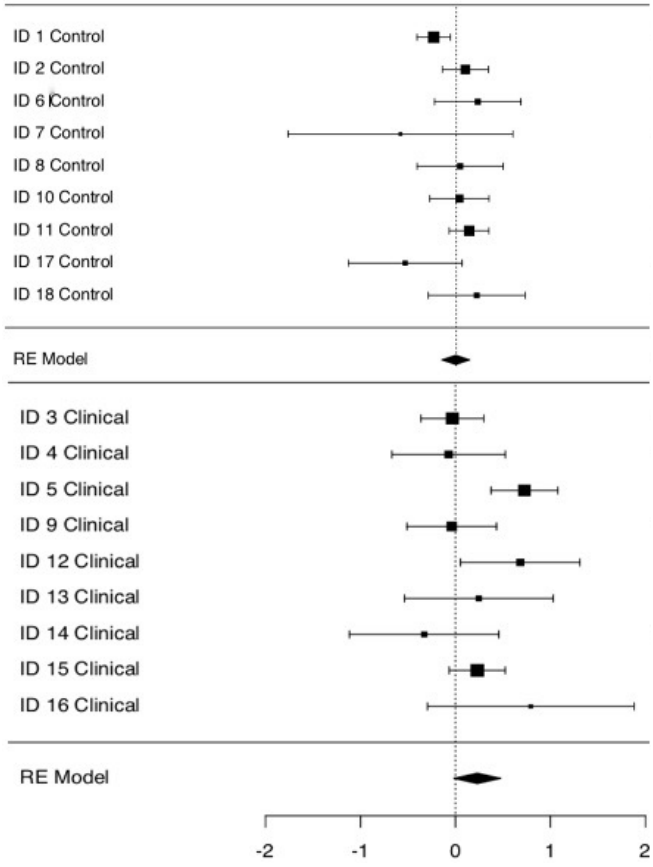


Focus Past

The within-subjects effects of *Time* (Jan-Jun 2019/ pre COVID, Jan-Jun 2020/COVID) on the use of the word category *Focus Past* are shown in Figure 3. There was no effect of *Time* on word use of *Focus Past*. A significant difference was found regarding the change of word use from 2019 to 2020 between the clinical and the control group (Z=2.23, p<0.01). There was a larger change in the use of the word category *Focus Past* from 2019 to 2020 in the clinical group relative to the control group, with a reduction of word use during COVID. The effects did not differ significantly within language.

Figure 3

Within-Subjects Effects Time/Focus Past



Mean Differences Blogposts About COVID/not About COVID

There was no mean difference in use of *I, Focus Future* and *Focus Past* in blogposts about COVID and blogposts not about COVID, controlling for language (see Table 4).

Table 4

Effect Topic Blogposts (COVID/not COVID)

Blogger	Mean	SD	n	Cohen's d	95% CI	p-value
			I-Use			
COVID	4.55	2.35	87	-0.03	-0.22, 0.15	.79
Not COVID	4.48	3.13	826			
			Focus Future			
COVID	1.88	1.13	87	0.01	-0.23, 0.25	.91
Not COVID	1.89	1.27	826			
			Focus Past			
COVID	3.60	1.89	87	0.11	-0.14, 0.35	.49
Not COVID	3.90	2.75	826			

Note. CI=Confidence Interval. A positive effect size indicates that blogposts about COVID included fewer words of the certain word category.

Discussion

The present study investigated the relationship of the COVID-19 pandemic as psychological stressor with the word use of bloggers, focusing on the word categories of first-person singular pronouns, future- and past orientation. Further, it was examined whether these variables are related differently between bloggers who suffer from a mental disorder and healthy control bloggers. It was hypothesized that blogpost published during the COVID-19 pandemic (January to June 2020) would include significantly more first-person singular pronouns, more past tense verbs and more future tense verbs, compared to blogposts published before COVID-19 emerged (January to June 2019). The second hypothesis was that the clinical group would show a significantly higher change in the use of these word categories, compared to the control group. Further, it was hypothesized that blogposts specifically dealing about COVID-19 would show a stronger difference in the use of those word categories, compared to blogposts about other topics.

The results revealed a main effect of patient status showing that the clinical group overall, used more first-person singular pronouns and fewer future tense verbs. Within-

subjects, there was no change in the use of the word categories *I*, *Focus Future* and *Focus Past* from pre COVID to COVID. Further, results indicated that the change in the use of the word category *Focus Past* from pre COVID to COVID was larger in the clinical group, relative to the control group, with a reduction of *Focus Past* during COVID. Additionally, results showed no difference in the use of the word categories comparing blogposts dealing specifically about COVID and blogposts not dealing about COVID.

In the following, these results are interpreted in terms of psychological health because as already elaborated, there is evidence that word-use is an indicator for mental health. For example, the already discussed study by Brockmeyer et al. (2015) investigating different clinical samples showed that I-use is related to symptoms of depression and anxiety. Related to that, the above elaborated study of Arntz et al. (2012) found that individuals with a personality disorder differed in their use of emotion words, verb tense and I-use from health controls.

Main Effect Patient Status

Interpreting the obtained results thus in terms of psychological health, the results showing a main effect of patient status are consistent with previous findings indicating that individuals with mental health problems differ from healthy individuals with regard to their word use. The significant difference in the use of first-person singular pronouns between the clinical- and the control group could be explained by excessive self-focus which is according to Harvey et al. (2004) an attentional mechanism being present in different forms of psychopathology, including depression and anxiety. This is in agreement with the self-awareness theory of reactive depression proposed by Pyszczynski and Greenberg (1987). As already elaborated, this theory proposes that depressed individuals are unable to disengage from a self-regulatory cycle focusing on losses and thereby develop an excessive self-focus. Hence, the results of the present study might also show that overall, the clinical group was more self-focused, measured by the use of first-person singular pronouns, which is consistent with previous literature showing that increased self-focus is a cognitive mechanism which is present in many mental disorders.

The main effect of patient status regarding the use of the word category *Focus Future* shows that individuals with mental health problems were less future oriented and thus, have a different time orientation compared to healthy controls. This is consistent with previous

studies showing the role of future orientation as protective factor for mental health. For example, Zheng et al. (2019) examined a sample of 369 participants investigating the effect of future orientation on stress related depression. Their results revealed that being future orientated helped individuals to better cope with stress and hence protected individuals from depression. Related to that, a study by Kim et al. (2019) found that future orientation mediated the effect between family stress and mental health problems in 638 African adolescents from low-income neighborhoods, meaning that participants who showed a higher future orientation showed less mental health problems when faced with familiar stressed compared to participants who were less future oriented. Thus, in terms of psychological health the results of the present study show that overall, individuals with mental health problems had a different time perspective, measured by the use of the word category *Focus Future* which is also in line with research indicating that future orientation has a mediating role between stress and mental health.

Change in Word Use from Pre COVID to COVID

No change was found within subjects in word use from pre COVID to COVID, which would mean in terms of psychological health that the COVID-19 pandemic did not affect mental health of participants. However, there are many studies showing the opposite, for example the already discussed results of Xiong et al. (2020) which showed that in the general population, higher symptoms of depression, anxiety and post-traumatic stress disorder were reported. This is supported by a review article of Vindegaard and Benros (2020) which included 43 studies examining the effect of the COVID-19 pandemic on mental health. Their systematic review showed that in the general population mental health decreased during the pandemic, that individuals with pre-existing mental disorders experienced increases in their psychiatric symptoms and that patients infected with COVID-19 showed increased symptoms of post-traumatic stress disorder and depression. Thus, there is much evidence for the negative impact COVID-19 has on mental health and the non-significant results regarding the change of word use from pre COVID to COVID in the present study might therefore show that word use is not a reliable indicator of mental health.

Nevertheless, there was still a difference found between groups in the change of *Focus Past,* with a greater reduction of past focus for the clinical group during COVID. In terms of psychological health that could mean that individuals with mental disorders and healthy

individuals are affected differently by the COVID-19 pandemic, reflected in the change of word use from pre COVID to COVID. This change in the word category *Focus Past*, as observed in the clinical group, might be related to the cognitive mechanism of rumination, being a repetitive, self-focused thinking pattern about *past* events and a typical symptom of depression (Hur et al., 2017). Hence, in terms of psychological health these results might also indicate that bloggers with mental disorders reduced more in their rumination from pre COVID to COVID relative to healthy bloggers. However, this result might also show again that word use is not a reliable measurement of mental health, as several studies showed that especially individuals with pre-existing mental disorders show an increase in their symptoms (Alonzi et al., 2020; Mazza et al., 2020), rather than a decrease as observed by the present results.

Topic of the Blog: COVID/not COVID

The results that word use did not differ with regard to the topic of the blog (COVID, not COVID) could mean in terms of psychological health that the preoccupation with the stressor (COVID) had no influence on mental health measured by word use of bloggers. This is based on the study of Cohn et al. (2004) who examined the relationship of word use and the September-11 attacks and found that the preoccupation with the attacks mediated the psychological response to this, reflected in the word use of individuals. Nevertheless, these null findings could also indicate again, that word use is not a reliable way of measuring mental health or not a good way of analyzing preoccupation with a stressor.

Reflection on Hypotheses

The first hypothesis that there would be a change in word use from pre COVID to COVID could not be confirmed, as there was overall no change in the use of these word categories from January until June 2019 to January until June 2020. These results contradict with those obtained by Cohn et al. (2004) and Pennebaker et al. (2002) who found that when experiencing stress and crises the use of different word categories, including verb tense and I-use changed. Further, these results are also in disagreement with recent studies published also examining the impact of COVID-19 on word use in different samples. For example, the study by Su et al. (2020) showed changes in word use in posts of the platform of Wuhan (*Weibo*) in 16 out of 39 word categories, including those analyzed in the present study. The posts from

the Italian platform (Twitter) showed significant changes in ten word categories, which were not the same as those analyzed in the present study. The already mentioned study Abdo et al. (2020) measured change in word use over 14 weeks analyzing tweets and also observed changes in several word categories. Thus, the present study obtained conflicting results compared to previous literature showing that word use changes in response to stressors, crises and also specifically in response to the COVID-19 pandemic. An explanation for the difference in results could be the long time period investigated in the present study. The studies presented above, analyzed fewer weeks for example the study of Su et al. (2020) focused on two weeks before and after the lockdown and the results obtained by Cohn et al. (2004) also showed that word use changed only on short term after the experience of a stressor. Concluding, there is evidence that word use changes only shortly after the experience of a stressor and an explanation for the difference in obtained results could be the large time frame analyzed in this study which could have led to not detecting a change.

The second hypothesis that there would be a difference in change between groups could only be confirmed partially, as only a difference between groups was found regarding the use of the word category of *Focus Past*. These results contradict with those obtained by Arntz et al. (2012) who compared essays written by individuals with a personality disorder and healthy controls. Their results showed amongst others a difference in the word use of verb tense and first-person singular use. Additionally, studies on the effects of COVID-19 on mental health found that individuals with pre-existing mental disorders show higher symptoms of amongst others depression and anxiety during the COVID-19 pandemic compared to healthy controls, so that it could have been expected that the clinical group would show a higher change in word use relative to the control group (Xiong et al., 2020; Alonzi et al, 2020). Nevertheless, a difference in the change between the clinical- and the control group was found regarding the change in *Focus Past* (which might reflect rumination of participants) although in the opposite direction than expected. This still shows that individuals with mental disorders and healthy individuals react differently to the COVID-19 pandemic, reflected in their word use of *Focus Past*. However, an alternative explanation for those results has to be considered which could be led back to the sample selected for the present study. Participants were assigned to the groups (clinical, control) based on whether they had published to be diagnosed with a mental disorder or not. Hence, it could not be controlled whether this assignment was based on the correct mental health status because it

was just based on the honesty of the participants on their blogs. This is something future studies should take into account by screening participants for mental health symptoms before the assignment to the groups.

The third hypothesis that blogposts dealing about COVID would include more words of the categories *I, Focus Future* and *Focus Past* compared to blogposts dealing about another topic could not be confirmed. This contradicts with the results obtained by Cohn et al. (2004) who found that the effects of the 9/11 terror attacks on mental health were stronger for individuals who were highly preoccupated with this stressor. An explanation of this difference in results could be the way preoccupation was measured. In contrast to the study by Cohn et al. (2004) who measured preoccupation with the proportional use of words related to the September-11 attacks, this study compared word use in blogposts dealing about COVID and blogposts not dealing about it as an indicator for preoccupation, as it was technically not possible to measure it with target words. Hence, it is questionable whether the measurement of this study really addressed the preoccupation of the individuals with COVID-19, which could have influenced the obtained results.

Next to that, a few other aspects of this study have to be considered. First, there is not much research available on word use as an indicator of psychological health, especially in relation with stress. Due to the novelty of the COVID-19 pandemic, there were also only a few studies published about how this virus affects mental health and only a very limited amount of research about the pandemic in relation with word use. Hence, the theoretical foundation on the present study could not be built on a sufficient number of studies specifically dealing about the topic of interest. However, this should not be concerned as a limitation because it underlies the great relevance this study had and the available opportunities for future research. Something else that has to be taken into account was the way data was decided to analyze in the present study. Analyses were conducted overall time frames however, an alternative way would have been to analyze change within months to see a clearer development in the use of word categories, which should be considered in the analyses of future research. A limitation of this study that has to be payed attention to is the limited number of blogs included, especially due to a deficiency of blogs written by bloggers who gave statements about having mental disorders. Further, as already mentioned not much information about the bloggers was available which limits the generalization of the results. Additionally, it could not be controlled whether the assignments to the groups (clinical,

control) really corresponded the bloggers' mental health status, as this was based on the bloggers' statements regarding their mental health on their blog.

Final Conclusion

Nevertheless, this study is of relevance. It was one of the first ones to focus on the influence of COVID-19 on mental health measuring the change in word use, which brings several implications. First, the presented results can help to understand and especially to further investigate the impacts of the COVID-19 pandemic on mental health reflected in word use of individuals. Further, the study used an implicit way to study mental health, compared to many studies using self-reports or clinical interviews. This reduces several biases and creates a more objective approach of measuring psychological health. Future studies should take the limitations of the present study into account and include a larger sample from which characteristics can be better controlled for, for example by screening them for mental health status. Studies should especially examine the development of mental health during the COVID-19 pandemic reflected in word use and consider analyzing changes in other word categories than those included in the present analysis. To increase the power of analyses, future studies could analyze text materials with same languages and clinical populations with the same mental disorders.

Concluding, this study was of great relevance as it was one of the first ones to investigate the change in word use from pre COVID to COVID comparing bloggers with mental health disorders and healthy controls. Results show that individuals with mental disorders and healthy controls use word categories differently. We did not observe a change in word use from pre COVID to COVID however, a difference in the change of past focus was found between groups suggesting that individuals with mental disorders and healthy controls react differently to the pandemic, reflected in word use. No difference in word use was found when bloggers wrote specifically about COVID compared to when they wrote about another topic. As the COVID-19 virus is present for over one year now and is proven to have a negative impact on mental health, this study underlies the importance of further examining the effects of the pandemic on psychological health and its relationship with word use.

References

Abdo, M. S., Alghonaim, A. S., & Essam, B. A. (2020). Public perception of COVID-19's global health crisis on Twitter until 14 weeks after the outbreak. *Digital Scholarship in the Humanities*. https://doi.org/10.1093/llc/fqaa037

Alonzi, S., La Torre, A., & Silverstein, M. W. (2020). The psychological impact of preexisting mental and physical health conditions during the COVID-19 pandemic. *Psychological Trauma: Theory, Research, Practice, and Policy*, *12*, 236-238. https://doi.org/10.1037/tra0000840

Arntz, A., Hawke, L. D., Bamelis, L., Spinhoven, P., & Molendijk, M. L. (2012). Changes in natural language use as an indicator of psychotherapeutic change in personality disorders. *Behaviour Research and Therapy*, *50*(3), 191-202. https://doi.org/10.1016/j.brat.2011.12.007

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis*. John Wiley & Sons.

Brockmeyer, T., Zimmermann, J., Kulessa, D., Hautzinger, M., Bents, H., Friederich, H., Herzog, W., & Backenstrass, M. (2015). Me, myself, and I: Self-referent word use as an indicator of self-focused attention in relation to depression and anxiety. *Frontiers in Psychology*, *6*, 1564. https://doi.org/10.3389/fpsyg.2015.01564

Clark, D. M. (2012). Cognitive behaviour therapy for anxiety disorders. *New Oxford Textbook of Psychiatry*, 1285-1298. https://doi.org/10.1093/med/9780199696758.003.0165

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge.

Cohn, M. A., Mehl, M. R., & Pennebaker, J. W. (2004). Linguistic markers of psychological change surrounding September 11, 2001. *Psychological Science*, *15*(10), 687-693. https://doi.org/10.1111/j.0956-7976.2004.00741.x

Harvey, A. G., Watkins, E., & Mansell, W. (2004). *Cognitive behavioural processes across psychological disorders: A transdiagnostic approach to research and treatment*. Oxford University Press, USA.

Hur, J., Heller, W., Kern, J. L., & Berenbaum, H. (2017). A Bi-factor approach to modeling the structure of worry and rumination. *Journal of Experimental Psychopathology*, *8*(3), 252-264. https://doi.org/10.5127/jep.057116

Junghaenel, D. U., Smyth, J. M., & Santner, L. (2008). Linguistic dimensions of psychopathology: A quantitative analysis. *Journal of Social and Clinical Psychology*, *27*(1), 36-55. https://doi.org/10.1521/jscp.2008.27.1.36

Kim, D. H., Bassett, S. M., So, S., & Voisin, D. R. (2019). Family stress and youth mental health problems: Self-efficacy and future orientation mediation. *American Journal of Orthopsychiatry*, *89*(2), 125-133. https://doi.org/10.1037/ort0000371

Mazza, M. G., De Lorenzo, R., Conte, C., Poletti, S., Vai, B., Bollettini, I., Melloni, E. M., Furlan, R., Ciceri, F., Rovere-Querini, P., & Benedetti, F. (2020). Anxiety and depression in COVID-19 survivors: Role of inflammatory and clinical predictors. *Brain, Behavior, and Immunity*, *89*, 594-600. https://doi.org/10.1016/j.bbi.2020.07.037

Mehl, M. R., Vazire, S., Ramirez-Esparza, N., Slatcher, R., & Pennebaker, J. W. (2007). Are women really more talkative than men? *Science*, *317*(5834), 82. https://doi.org/10.1126/science.1139940

Monroe, S. M., & Slavich, G. M. (2016). Psychological Stressors: Overview. In G. Fink (Ed.), *Stress: Concepts, cognition, emotion, and behavior: Handbook of stress series* (pp. 109-115). Academic Press.

Pennebaker, J. W., Booth, R. J., Boyd, R. L., & Francis, M. E. (2015). *Linguistic inquiry and word count: LIWC2015*. Austin, TX:Pennebaker Conglomerates.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. Austin, TX: University of Texas at Austin.

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology*, *77*(6), 1296-1312. https://doi.org/10.1037/0022-3514.77.6.1296

Pennebaker, J. W., & Lay, T. C. (2002). Language use and personality during crises: Analyses of mayor Rudolph Giuliani's press conferences. *Journal of Research in Personality*, *36*(3), 271-282. https://doi.org/10.1006/jrpe.2002.2349

Pyszczynski, T., & Greenberg, J. (1987). Self-regulatory perseveration and the depressive self-focusing style: A self-awareness theory of reactive depression. *Psychological Bulletin*, *102*(1), 122-138. https://doi.org/10.1037/0033-2909.102.1.122

Su, Y., Xue, J., Liu, X., Wu, P., Chen, J., Chen, C., Liu, T., Gong, W., & Zhu, T. (2020). Examining the impact of COVID-19 lockdown in Wuhan and Lombardy: A

psycholinguistic analysis on Weibo and Twitter. *International Journal of Environmental Research and Public Health*, *17*(12), 4552. https://doi.org/10.3390/ijerph17124552

Vindegaard, N., & Benros, M. E. (2020). COVID-19 pandemic and mental health consequences: Systematic review of the current evidence. *Brain, Behavior, and Immunity*, *89*, 531-542. https://doi.org/10.1016/j.bbi.2020.05.048

World Health Organization. (2020, March 12). *WHO erklärt COVID-19-Ausbruch zur Pandemie*. https://www.euro.who.int/de. Retrieved March 29, 2021, from https://www.euro.who.int/de/health-topics/health-emergencies/coronavirus-covid-19/news/news/2020/3/who-announces-covid-19-outbreak-a-pandemic

Xiong, J., Lipsitz, O., Nasri, F., Lui, L. M., Gill, H., Phan, L., Chen-Li, D., Iacobucci, M., Ho, R., Majeed, A., & Mclntyre, R. S. (2020). Impact of COVID-19 pandemic on mental health in the general population: A systematic review. *Journal of Affective Disorders*, *277*, 55-64. https://doi.org/10.1016/j.jad.2020.08.001

Zheng, L., Lippke, S., Chen, Y., Li, D., & Gan, Y. (2019). Future orientation buffers depression in daily and specific stress. *PsyCh Journal*, *8*(3), 342-352. https://doi.org/10.1002/pchj.283

Appendix

Descriptive Statistics Word Categories pre COVID and COVID

	Pre COVID							COVID						
Blogger	**N**	**M I**	**SD I**	**M future**	**SD future**	**M past**	**SD past**	**N**	**M I**	**SD I**	**M future**	**SD future**	**M past**	**SD past**
Clinical Group														
ID 3	61	6.27	3.06	2.61	1.33	5.11	1.64	83	5.62	3.06	2.63	1.3	5.17	1.89
ID 4	16	6.73	2.67	3.17	1.19	5.47	1.92	33	5.71	2.75	2.75	0.93	5.63	2.3
ID 5	70	4.92	2.11	1.33	0.46	4.66	1.48	63	4.44	1.74	1.41	0.5	3.61	1.39
ID 9	39	3.32	2.32	1.11	0.75	3.07	1.95	31	3.93	2.16	1.11	0.52	3.15	2.06
ID 12	24	7.03	2.53	1.24	0.95	3.55	1.79	18	5.67	3.04	1.6	1.2	2.35	1.64
ID 13	10	6.29	2.49	1.04	0.67	4.69	2.2	17	5.4	1.7	1.27	0.65	4.09	2.46
ID 14	29	6.06	2.16	0.91	0.56	3.5	1.78	8	5.89	2.1	0.87	0.87	4.12	2.07
ID 15	97	5.68	2.6	0.81	0.57	2.49	1.47	82	6.31	2.88	0.88	0.48	2.15	1.48
ID 16	22	6.93	2.21	0.71	0.52	3.88	1.77	4	5.28	3.67	0.68	0.19	2.49	1.08
Control Group														
ID 1	169	6.34	2.03	2.57	1.15	5.63	2.08	204	5.93	2.56	2.68	1.48	6.25	2.33
ID 2	99	6.12	2.2	2.87	1.18	6.02	1.99	92	5.45	1.88	2.73	0.96	5.78	2.29
ID 6	26	3.41	2.21	0.68	0.44	2.4	1.22	29	3.98	2.82	0.94	0.47	2.06	1.31
ID 7	7	5.05	2.14	1.07	0.47	3.25	1.6	3	4.88	3.71	0.89	0.23	5.14	4.11
ID 8	36	3.4	1.16	1.14	0.7	1.69	0.81	22	3.8	1.56	1.21	0.56	1.65	0.85
ID 10	72	3.1	2.43	1.28	0.76	2.79	1.88	48	4.22	2.79	1.19	0.79	2.72	1.72
ID 11	118	0.56	1.04	1.67	0.95	1.2	0.93	141	0.17	0.65	1.6	0.94	1.07	0.72
ID 17	18	0.4	0.67	0.66	0.54	0.87	0.52	15	1.11	1.79	0.85	0.46	1.29	0.78
ID 18	24	3.91	1.27	0.74	0.41	2.3	0.94	20	4.65	1.47	0.64	0.38	2.08	0.75